

(19)



Europäisches Patentamt  
European Patent Office  
Office européen des brevets



(11)

**EP 0 713 304 A3**

(12)

**EUROPEAN PATENT APPLICATION**

(88) Date of publication A3:  
03.11.1999 Bulletin 1999/44

(51) Int Cl.<sup>6</sup>: H04L 12/56, H04Q 11/04

(43) Date of publication A2:  
22.05.1996 Bulletin 1996/21

(21) Application number: 95308196.5

(22) Date of filing: 15.11.1995

(84) Designated Contracting States:  
DE FR GB

(72) Inventor: Lyles, Joseph B.  
Mountain View, California 94043 (US)

(30) Priority: 15.11.1994 US 340493

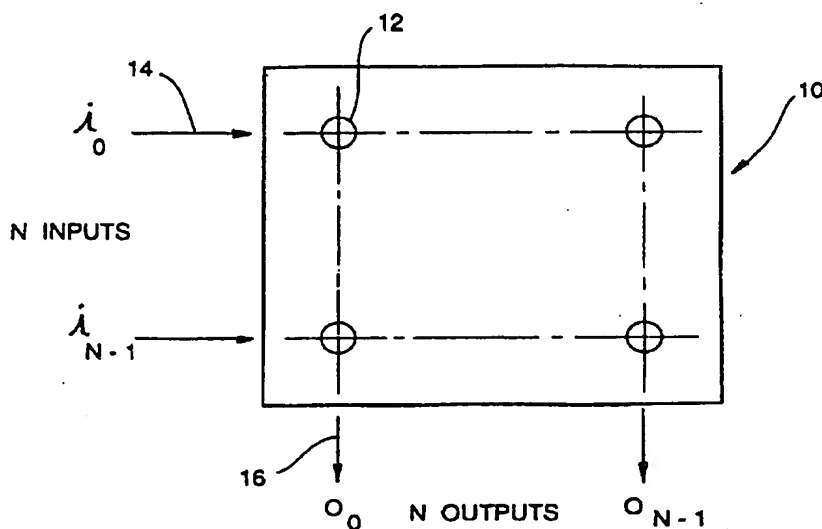
(74) Representative: Grünecker, Kinkeldey,  
Stockmalr & Schwanhäusser Anwaltssozietät  
Maximilianstrasse 58  
80538 München (DE)

(71) Applicant: XEROX CORPORATION  
Rochester, New York 14644 (US)

(54) **ATM crossbar switch fabric**

(57) An expandable, self-routing switching fabric routes data packets in an ATM switch (10). A crosspoint array of state machines (Fig. 2) disposed at internal intersections of input ports (14) and output ports have predetermined identities (Fig. 3) representative of the output ports (16) and include a comparator for identifying an address header of each data packet with the state machine identity. When a match occurs (Fig. 2), the

state machine is set active/connected and when a non-match occurs, the state machine is set not-active/disconnected. A circuit through the fabric is set by connection between an input port (14) through an active/connected state machine to a desired output port (16). The predetermined identity of each state machine (Fig. 2) comprises a hard wired bit sequence (Fig. 3) common to the column including the state machines.

**FIG. 1**

EP 0 713 304 A3

## Description

This invention relates to high-speed, broadband communication networks and, more particularly, to a self routing crossbar switch suitable for communicating data packets in an ATM switch.

High-speed, broad-band communication networks include systems for routing data packets from input sources to output sources. US-A-s 5,305,311 and 5,327,420 disclose such switching networks. The basic architecture of such prior art systems include an Asynchronous Transfer Mode (ATM) Local Area Network (LAN) using a VLSI-based "Batcher/banyan" switching fabric of the type which has been developed by AT&T/Bellcore. Batcher/banyan switching technology will scale to switches larger than 256 ports, and link rates exceeding 1 gigabit/sec and these switches can be interconnected to form a larger hierarchical network for supporting larger user communities.

An advantage of such Batcher/banyan switch fabrics is that they are "self routing". That is, data packets are routed through the switching fabric on the basis of local decisions only, there is no global "chokepoint". For each stage in the network, routing can be determined with only the information contained in the data packets entering that stage of the switch; that is, with locally available information. Such networks are also advantageous when they are "non-blocking". In any switching cycle, the network is capable of generating an arbitrary permutation, or a partial permutation, of its inputs. More simply stated, a nonblocking switching fabric is able to generate all possible one to one input/output mappings in a single switching cycle.

Ideally such a switching fabric provides parallelism whereby a single N port Batcher sorting network can feed some number k (typically two or three) banyan routing switches that are connected so that as many as k-cells can be routed to the same number of output ports during the same cell switching cycle without losing any of the data packets. However, in order to overcome the problem that occurs when more than one data packet is found to be destined to the same output port during the single switch cycle, a "reservation ring" is provided for fairly adjudicating the contention among such conflicting data packets, while permitting at most k of them to be presented to the switch in any cycle. Reservation rings resolve output contentions among conflicting data packets, while implementing "fair" access to the output ports of the switching fabric.

The first bits of a data packet contain the address to which the packet should be routed. Such "first bits" are typically referred to as the "header", and identify the path through the switching fabric, the source and destination of which are input and output ports. For a particular switch, the input and output ports associated with a given virtual circuit are defined at connection establishment. A virtual circuit is thus the ultimate source and ultimate destination, and all the intermediate switches

therebetween defining the path through the fabric. Accordingly, the contents of the header are defined by the virtual circuit as the path through the particular switching fabric. Switching fabric such as Batcher/banyans allow many possible paths between any given input and output pair and present a problem which is sought to be overcome by the present invention. The problem is that it is very difficult to diagnose failures in a Batcher/banyan switch fabric due to the multiple possible paths. A failure is difficult to replicate because of the many possible paths.

The present invention contemplates a new and improved switching fabric to replace the Batcher/banyan network of prior art systems, provide easier diagnosis of failures in the fabric than in the Batcher/banyan network, is simple in design, economical to manufacture and which provides better scaling properties than a Batcher/banyan for the size of the switching fabrics that the subject invention is applicable to be implemented as in a commercial embodiment.

In accordance with the present invention, there is provided an expandable, self routing switching fabric for routing data packets in an ATM switch comprising: a plurality of input ports and a plurality of output ports for routing the packets from input sources coupled to said input ports to output destinations coupled to said output ports on circuits through the fabric specified by the packets; and a crosspoint array of state machines disposed at internal intersections of the input ports and the output ports, wherein each state machine has a predetermined identity representative of a one of the output ports and includes a means for comparing the identity to an address header of each data packet and for setting each state machine active/connected when the comparing determines a match between the identity and the address header and for setting each state machine not active/disconnected when the comparing determines a non-match between the identity and the address header, whereby a selected circuit through the fabric to a desired output port comprises an active/connected state machines in connection from a one of the input ports to the desired output port.

The switching fabric of the invention comprises a crosspoint array wherein the input ports comprise a plurality of rows and the output ports comprise a plurality of columns. The selected circuit is the input row and the output column through the active/connected state machine defined by the desired output port.

In accordance with a more limited aspect of the subject invention, the predetermined identity of each state machine comprises a hard-wired bit sequence common to the column including the active/connected state machines defined by the desired output port.

One benefit obtained by the present invention is an ATM switch fabric which is simple in design and thus easier to diagnose for failures than prior art Batcher/banyan crossbar switches.

Another benefit obtained by the present invention

is a switching fabric which has better scaling properties for small scaling values.

The invention may take physical form in certain parts and arrangements of parts, the preferred embodiments of which will be described in this specification and illustrated in the accompanying drawings which form a part hereof, and wherein:

FIGURE 1 is a diagrammatic illustration of a crossbar chip, of size  $N \times N$ , as intended as part of the present invention;

FIGURE 2 defines the possible states of each state machine at the crosspoints of the chip of FIGURE 1; FIGURE 3 illustrates a tabular sequence for comparing an address identity of an output port or column against header information of the cell packet; FIGURE 4 illustrates an array of the crossbar chips of FIGURE 1 connected to form a switching fabric  $2N \times 2N$ ;

FIGURE 5 shows an expanded switching fabric of the crossbar chips of FIGURE 1 having a speed-up of 2;

FIGURE 6 is a tabular illustration of the sequence for comparing header information against column identity in an 8 by 8 switching fabric with a speed-up of 2;

FIGURE 7 shows a standard form synchronous state machine such as can be employed in the crosspoint array of the subject invention; and, FIGURE 8 is a tabular definition of the state machine of FIGURE 7.

Referring now to the drawings, the various FIGURES show a self routing crossbar switch suitable for use as a switching fabric in an ATM switch. The invention is particularly applicable to replace the Batcher/banyan network in a system such as is shown in US-A-5,327,420.

The basic architecture of such a system is directed to high-speed, parallel transmission and switching of small, fixed sized data packets, called "cells". The first bits of each data packet, called a "header", contain an address indicative of an output port to which the packet should be routed.

FIGURE 1 illustrates a crossbar chip 10 representing a fundamental element of the fabric, of size  $N \times N$ , where  $N$  is a power of 2, which forms an  $N$  by  $N$  switching fabric. The fabric comprises a crosspoint array in which the crosspoints 12 are located at intersections of an input row 14 and output column 16. The inputs are illustrated as ranging from  $I_0$  to  $I_{n-1}$  and similarly, the outputs are represented as ranging from  $O_0$  to  $O_{n-1}$ . It is, of course, the overall objective of the invention to transmit a cell from a one of the input ports through the fabric to a desired output port in a quick and reliable manner.

For such an ATM switch, each cell will be presented bits serially, most significant bit first, and will have the header pre-pended to the cell. Presentation of the cells

will occur synchronously at all inputs 14. Thus, the self routing crossbar chip can compare all cell headers in lock step. If a given input is inactive during a cell time, it still presents a dummy cell, with the leading bit set to "0" to denote "inactive".

The address is in a form to identify the output column 16 to which the cell should be routed out within the  $N \times N$  chip. Each chip 10 has  $N^2$  internal intersections at the crosspoints between the rows and the columns. At each of the intersections is disposed a state machine having an operating character as shown in FIGURE 2. The most significant bit of a field of the cell is transmitted first on the rows 14. At each state machine at the crosspoints 12, each bit of the address is compared against a hard wired bit sequence common to the column of the intersection. The sequence comprises a predetermined identity representative of the output port common to all state machines of the particular column. The array is completely regular if the address generated for each state machine is common to the column and not present at each state machine. If at any step in the sequential comparison the state machine makes a determination of "not-equal", then the state machine goes "not-active/disconnected". At the end of the addresses, only the active state machines can connect a particular row to the desired column output. In the case of maximum operating efficiency of the fabric,  $N$  cells, each on a one of the input rows, can be individually connected to the  $N$  output ports for simultaneous transmission through the switching fabric. In order to avoid conflicting cells seeking output to the same port, a reservation ring (not shown) insures that only one row will address a desired column, so that only one input will drive a given output.

With particular reference to FIGURE 3, a more detailed illustration of the subject invention is shown wherein a 4 by 4 switching fabric with columns numbered "00" through "11" are shown. It can be seen that the address comparing occurs in three time steps. During time step 0, the most significant bits are set to "1" to indicate that the cell is active. If the cell were inactive, the most significant bit would be set to "0". During time steps 1 and 2, the column is selected within the chip. It can be seen that the first column is identified by the sequence "00" and the fourth column is identified by the sequence "11". Thus, all rows will be compared against the sequence "100" for column "00" and against sequence "111" for column "11".

It is an important feature of the invention that the switching fabric 10 can be generalized to both larger fabrics than are possible with a single chip and to fabrics where there is a speed-up of  $k$  from the inputs to the outputs (analogous to the speed-up of  $k$  achievable with  $k$  banyans in a Batcher/banyan network).

FIGURE 4 shows an array 20 of four of the crossbar chips 10, each  $N$  by  $N$  in size, where  $N$  is a power of 2, connected to form a switching fabric of  $2N \times 2N$ . The chips 10 are connected in 2 columns 22, 24 of two  $N$  by  $N$  crossbars; each crossbar column is given an identity

by means of pins tied to ground or VCC. Thus, in addition to each individual output port having an identity, the two columns each have an identity as well. Also, as with Batcher/banyan networks, a speed-up of  $k$  can be achieved by multiplying the number of columns by  $k$ . For example, FIGURE 5 shows such a circuit with  $2N$  inputs and  $2 \times 2 \times N$  outputs for a speed-up value of  $k = 2$ . The identity of the desired output port is represented by a single address in the form: [cell active/inactive, value of  $k$ , chip column, column within an  $N \times N$  chip], so that each input to the switching fabric identifies by its address the desired output port.

With reference to FIGURES 5 and 6, the above address form can be specifically illustrated. FIGURE 5 shows an 8 by 8 switching fabric with a speed-up of 2. There are sixteen columns spread across four columns of 4 by 4 chips, with the individual columns numbered "0000" through "1111". In FIGURE 6, the sequence for comparing cell address against the identity of each column is shown. Thus, all rows will be compared against the sequence "10000" for column "0000" and against "11111" for column "1111". In time step 0, a comparison occurs to determine whether the cell is active or not active. In time step 1, the  $k$  value comparison occurs. In time step 2, the chip column comparison occurs and in time steps 3 and 4, the select column within each of the chips comparison occurs.

With reference to FIGURES 7 and 8, one possible hardware permutation of a state machine having a state diagram as illustrated in FIGURE 2 is shown. A state machine as shown in FIGURE 7 will be disposed at each crosspoint 12 (FIGURE 1) in the switching fabric 10. FIGURE 8 provides a detailed definition of function for the state machine in FIGURE 7. This implementation requires approximately twelve gate equivalents. A 128 by 128 self routing crossbar switch could be easily implementable using standard cell technology.

A particular advantage of the present invention is that it is easier to diagnose failures in the switching fabric than in prior known Batcher/banyan networks. In particular, such prior art systems required each switching point to examine one or more of the cell header address bits to decide how to subsequently route the rest of the packet. The resulting virtual circuit was complex and it was hard to identify where failure occurred. The present invention overcomes this problem by providing a simplified system wherein a failure can be located at the state machine in the column of the fabric common to the desired output port.

Another advantage of the present invention is its ability to be readily expandable with better scaling properties. For example, a  $2N$  Batcher/banyan network requires ten chips as opposed to the eight chips for a comparable self routing crossbar switch manufactured in accordance with the subject invention. Although it will be appreciated by one of ordinary skill in the art that for large  $N$  switching fabrics a Batcher/banyan may scale better than the subject invention.

#### Claims

1. An expandable, self routing switching fabric for routing data packets in an ATM switch comprising:

a plurality of input ports (14) and a plurality of output ports (16) for routing the packets from input sources coupled to said input ports to output destinations coupled to said output ports on circuits through the fabric specified by the packets; and,

a crosspoint array of state machines (Fig. 2) disposed at internal intersections of the input ports (14) and the output ports (16), wherein each state machine has a predetermined identity representative of a one of the output ports and includes a means for comparing the identity to an address header of each data packet and for setting each state machine active/connected when the comparing determines a match between the identity and the address header and for setting each state machine not active/disconnected when the comparing determines a non-match between the identity and the address header,

whereby a selected circuit through the fabric to a desired output port (16) comprises an active/connected state machines (Fig. 2) in connection from a one of the input ports (14) to the desired output port.

2. The switching fabric as defined in claim 1 wherein the input ports (14) comprise a plurality of rows of the crosspoint array and the output ports (16) comprise a plurality of columns of the crosspoint array and wherein the selected circuit comprises a one of the columns of active/connected state machines (Fig. 2) defined by the desired output port.
3. The switching fabric as defined in claim 2 wherein the predetermined identity of each state machine (Fig. 2) comprises a hardwired bit sequence (Fig. 3) common to the one of the columns including said each state machine.
4. The switching fabric as defined in claim 1, 2 or 3 wherein the switching fabric comprises an  $N$  plurality of input ports (14) and a  $kN$  plurality of output ports (16) where  $k$  represents a speed up of the switching fabric from the input ports to the output ports.
5. The switching fabric as defined in claim 4 wherein the address header comprises a form representative of data packet active/inactive, a value of  $k$ , a chip column and column within an  $N \times N$  chip, wherein the switching fabric is comprised of a plurality of  $N \times N$  chips.

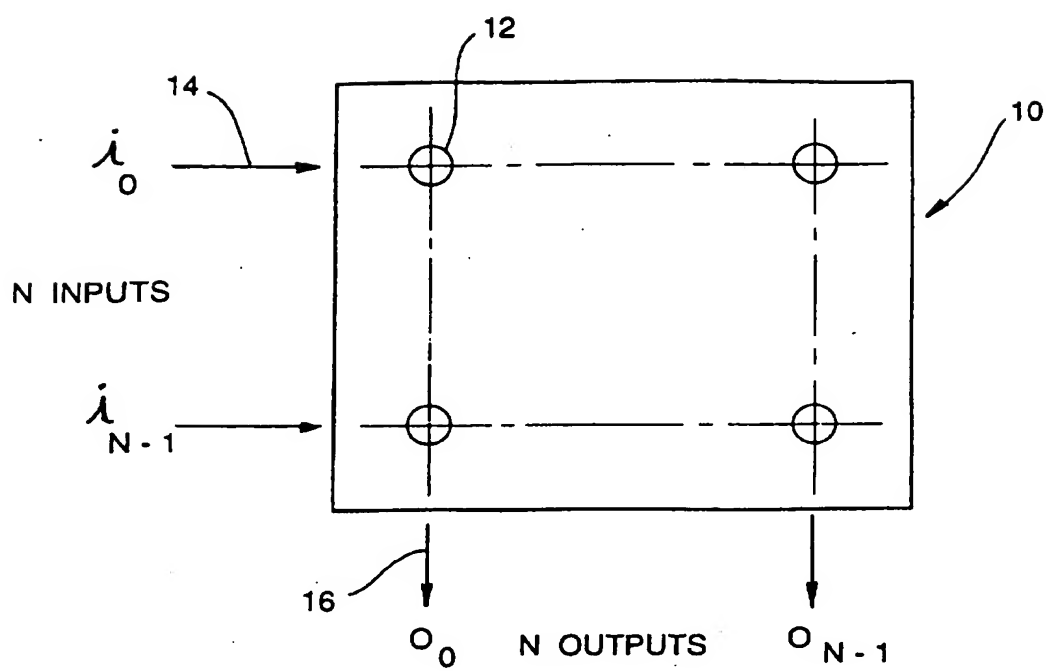


FIG. 1

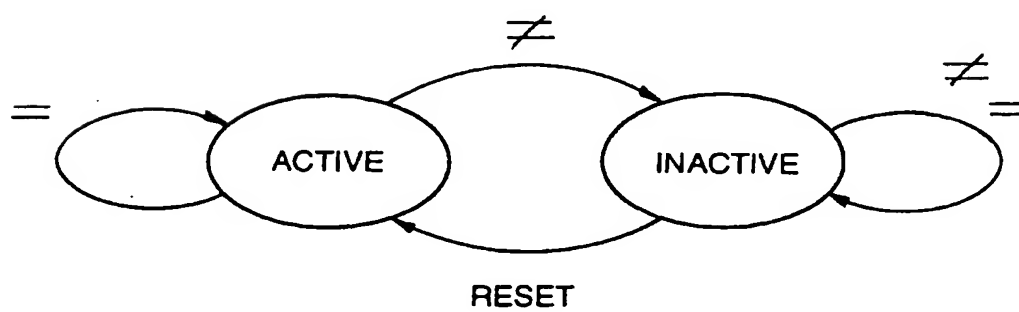


FIG. 2

TIME STEP	COLUMN			
	00	01	10	11
	0	1	1	1
	1	0	0	1
	2	0	1	1

FIG. 3

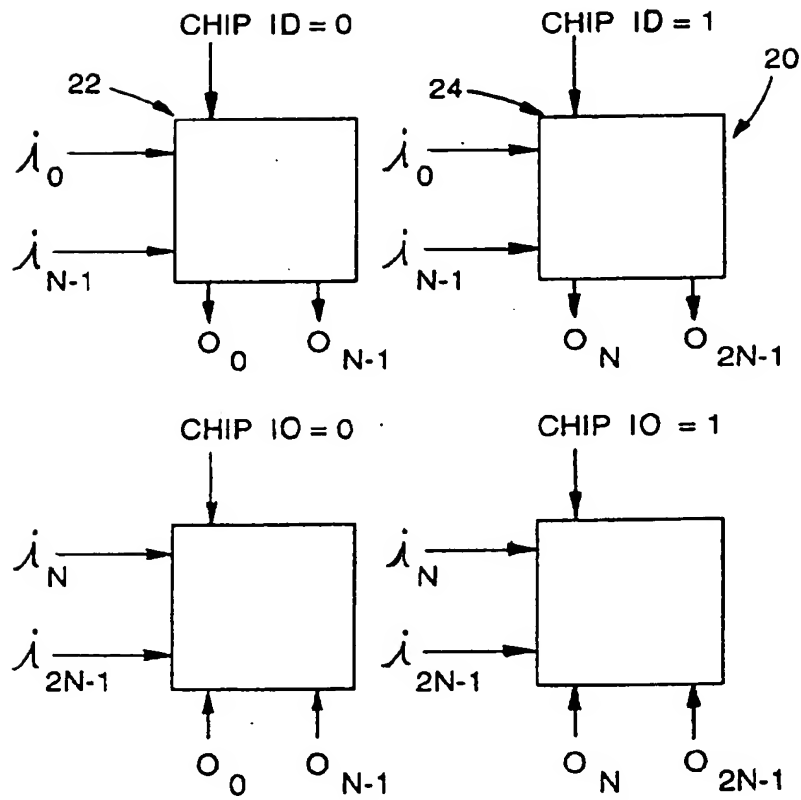
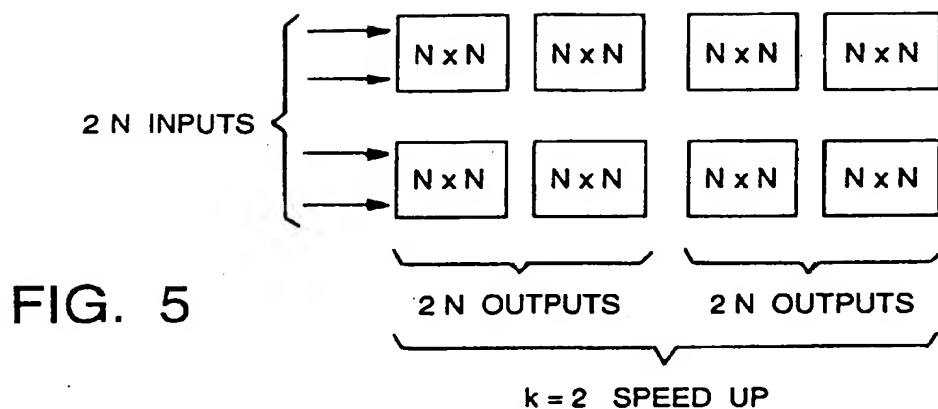


FIG. 4



	K VALUE = 0								K VALUE = 1							
	CHIP COL. 0				CHIP COL. 1				CHIP COL. 2				CHIP COL. 3			
OUTPUT NUMBER	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
2	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
3	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
4	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1

TIME STEP 0 COMPARES ACTIVE BIT

TIME STEP 1 COMPARES k VALUE

TIME STEP 2 COMPARES CHIP COLUMN

TIME STEPS 3 & 4 COMPARE & SELECT COLUMNS WITHIN A CHIP

**FIG. 6**

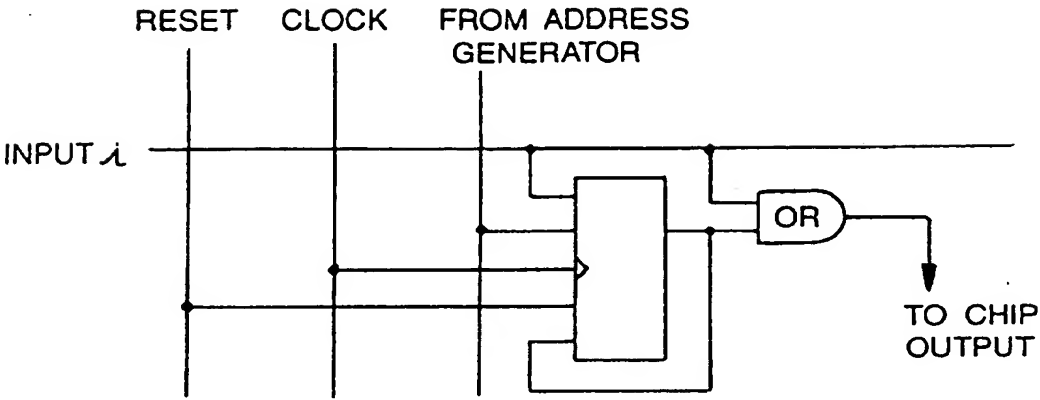


FIG. 7

INPUT	ADDRESS	CURRENT STATE	RESET	NEXT STATE
0	0	0	0	0
0	1	0	0	0
1	0	0	0	0
1	1	0	0	0
0	0	1	0	1
0	1	1	0	0
1	0	1	0	0
1	1	1	0	1
0	0	0	1	1
0	1	0	1	0
1	0	0	1	0
1	1	0	1	1
0	0	1	1	1
0	1	1	1	0
1	0	1	1	0
1	1	1	1	1

FIG. 8





European Patent  
Office

# EUROPEAN SEARCH REPORT

Application Number  
EP 95 30 8196

DOCUMENTS CONSIDERED TO BE RELEVANT			
Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int.Cl.6)
X	CLOONAN T J ET AL: "SELF-ROUTING CROSSBAR PACKET SWITCH EMPLOYING FREE-SPACE OPTICS FOR CHIP-TO-CHIP INTERCONNECTIONS" APPLIED OPTICS, vol. 30, no. 26, 10 September 1991 (1991-09-10), pages 3721-3733, XP000226959 ISSN: 0003-6935	1-3	H04L12/56 H04Q11/04
Y	* figures 3-6,8 * * page 3723, column 2, line 6 - line 29 * * page 3723, column 2, line 34 - line 38 * * page 3724, column 1, line 15 - line 33 * * page 3724, column 2, line 17 - page 3725, column 1, line 14 * * page 3732, column 1, paragraph V *	4	
X	MASAO AKATA ET AL: "A 250MB/S 32 X 32 CMOS CROSSPOINT LSI FOR ATM SWITCHING SYSTEMS" IEEE INTERNATIONAL SOLID STATE CIRCUITS CONFERENCE, 14 February 1990 (1990-02-14), pages 30-31, 257, XP000201818 ISSN: 0193-6530 * figures 2,4 * * page 30, column 1, line 13 - line 17 * * page 30, column 1, line 31 - line 48 * --- -/--	1-3	TECHNICAL FIELDS SEARCHED (Int.Cl.6) H04Q H04L
The present search report has been drawn up for all claims			
Place of search THE HAGUE		Date of completion of the search 13 September 1999	Examiner Lamadie, S
CATEGORY OF CITED DOCUMENTS X : particularly relevant if taken alone Y : particularly relevant if combined with another document of the same category A : technological background O : non-written disclosure P : intermediate document		T : theory or principle underlying the invention E : earlier patent document, but published on, or after the filing date D : document cited in the application L : document cited for other reasons & : member of the same patent family, corresponding document	

EPO FORM 1503 03.92 (P/MC01)



European Patent  
Office

# EUROPEAN SEARCH REPORT

Application Number  
EP 95 30 8196

DOCUMENTS CONSIDERED TO BE RELEVANT			
Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int.Cl.6)
Y	CHAO H J: "A RECURSIVE MODULAR TERABIT/SECOND ATM SWITCH" IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, vol. 9, no. 8, 1 October 1991 (1991-10-01), pages 1161-1172, XP000267569 ISSN: 0733-8716	4	
A	* figures 2,6-10 * * page 1165, column 1, line 25 - column 2, line 12 * * page 1167, column 1, line 17 - line 30 * * page 1170, column 1, line 33 - line 40 *	1,5	
			TECHNICAL FIELDS SEARCHED (Int.Cl.6)
The present search report has been drawn up for all claims			
Place of search THE HAGUE		Date of completion of the search 13 September 1999	Examiner Lamadie, S
<p>CATEGORY OF CITED DOCUMENTS</p> <p>X : particularly relevant if taken alone Y : particularly relevant if combined with another document of the same category A : technological background O : non-written disclosure P : intermediate document</p> <p>T : theory or principle underlying the invention E : earlier patent document, but published on, or after the filing date D : document cited in the application L : document cited for other reasons &amp; : member of the same patent family, corresponding document</p>			

EPO FORM 1503 (3.92) (P4001)